

ON NUMERICAL METHODS FOR CLASSIFYING RELEVÉS COLLECTED IN BRAUN-BLANQUET PHYTOSOCIOLOGICAL SURVEYS

B. M. CAMPBELL AND E. J. MOLL

(Department of Botany, University of Cape Town)

ABSTRACT

The case for using a numerical technique for classifying relevés collected in Braun-Blanquet phytosociological surveys is presented. It is suggested that the method of group-average sorting based on the Canberra similarity measure is a suitable technique. Information loss or distortion is minimal, and the success of the similarity measure is independent of the nature of the raw data. This computer-based method of classifying relevés provides an ecologically interpretable hierarchical classification. Using this classification and a further computer programme for subjectively re-arranging species on the phytosociological table it is possible to effortlessly construct the final table.

UITTREKSEL

NUMERIESE METODES VIR DIE KLASSIFIKASIE VAN RELEVÉS VERSAMEL IN 'N BRAUN-BLANQUET FITOSOSIOLOGIESE OPNAME

Regverdiging vir die gebruik van numeriese tegnieke om die relevés in 'n Braun-Blanquet fitososiologiese opname word aangebied. Dit word voorgestel dat die metode van groep-gemiddelde sortering gebaseer op die Canberra gelykvormheidsmaat 'n geskikte tegniek is. Inligtings verlies of verwringing is minimaal en die sukses van die gelykvormigheidsmaat is onafhanklik van die soort onverwerkte data. Hierdie rekenaar-gebaseerde metode om relevés te klassifiseer voorsien 'n ekologiese vertolkbare hiërgiesse klassifikasie. Deur hierdie klassifigasie te gebruik asook 'n verdere rekenaarprogram om die soorte subjektief te herrangskik op die fitososiologiese tabel is dit moontlik om sonder moeite die finale tabel saam te stel.

INTRODUCTION

In Southern Africa the Braun-Blanquet or Zurich-Montpellier method of vegetation survey is becoming increasingly popular (Werger, 1974a). In most surveys the Braun-Blanquet table method (Werger, 1974a; Shimwell, 1971) has been used to construct the final phytosociological table. By this method relevé similarity is visually assessed and the resulting vegetation classification is therefore entirely subjective. Even though the results of this method are often very similar to those obtained by numerical methods (e.g. Spaty et Siegmund, 1973, in Werger, 1974a) this method remains subjective and different workers dealing with the same data are likely to produce different classifications, especially at the lower hierarchical level where there are only subtle differences between relevés and groups of relevés.

Accepted for publication 8th September, 1975.

This paper is primarily aimed, not at statistical ecologists, but at phytosociologists who wish to obtain numerically-based vegetation classifications. We suggest a numerical method which will provide a classification of relevés for general purposes and which requires no subjective refinement.

DISCUSSION

There are a large number of numerical techniques which could be used to obtain numerically-based vegetation classifications (Williams, 1971) and a number of techniques have been specifically developed for constructing phytosociological tables (Češka and Roemer, 1971; Schmid and Kuhn, 1970 in Werger, 1974a). In two recent papers (Coetzee, 1974; Coetzee and Werger, 1973) the Braun-Blanquet table method was argued to be superior to the numerical methods tested. Both the numerical methods used by these workers have undesirable properties. This is the case with many numerical methods, the efficiency of the method often depending on the nature of the raw data.

In this paper we have considered only those methods which have proved popular or have been developed in South Africa. The method we suggest appears suitable irrespective of the raw data matrix. For accounts of the many methods available reference should be made to Gower (1967), Williams (1971) and Williams, Lance, Webb and Tracey (1973).

Association analysis

One method which should be considered is Association Analysis (Williams and Lambert, 1959; 1961). This method has been widely used in synecological studies (Werger, 1974a) but has had limited success and has, therefore, fallen into general disuse. Failure can be attributed to Association Analysis being a monothetic method, *i.e.* a method in which the groupings in the classification may be defined by a single species. As Coetzee (1974) points out, no species is usually 100% constant in a group of distinctly related relevés, and the chance absence of a defining species from a relevé will result in misclassifications.

If a numerical method is to be useful it should be polythetic, the groupings being based on overall similarity.

Some polythetic methods

Most polythetic methods consist of two steps. The first is the computation of measures of similarity between all relevés to be classified on the basis of their total floristic composition, and secondly the hierarchical arrangement of relevés by sorting through the similarity matrix.

1. Similarity coefficients

The choice of a similarity coefficient requires special consideration of two possible properties of the coefficient (Williams *et al.*, 1973). Firstly, many

coefficients are abundance weighted. The more abundant species may, therefore, dominate the analysis and potential information from less abundant species may be lost. Such coefficients are undesirable as they require an intuitive assessment of the data prior to analysis as standardisation may be necessary to ensure the success of the coefficient; for example square-root or logarithmic transformation (Field, 1971; Walker, 1974).

However, abundance-weighted coefficients appear not too problematical with Braun-Blanquet phytosociological data, as the cover-abundance values are semi-logarithmically transformed relative to the actual cover. For example species with 76 to 100% cover having a cover-abundance value of 5 while species with 1 to 5% cover being considered only 5 $\times$ less 'important', *i.e.* having a value of 1. For this reason the results of a classification using the abundance-weighted Bray and Curtis measure* are almost identical to that given by the Canberra measure which is not abundance weighted (Fig. 1.). The groups recognised in the classifications are easily ecologically explainable.† Although the Bray and Curtis measure has proved successful, its abundance weighting property will be undesirable for some data matrices. The Euclidean metric is excessively sensitive to relative abundance (Williams *et al.*, 1973) and its success is, therefore, dependent on the nature of the raw data.

The second property which requires consideration is whether the similarity measure excludes double zero matches, *i.e.* joint absences of given species. That these matches are excluded is desirable for, as Field (1969) suggests, 'no marine ecologist would consider that the intertidal and abyssal faunas were similar because both lacked the species found on the continental shelf'. Measures which include double zero matches are, therefore, of restricted use, their success depending on the nature of the raw data. They will only be successful with raw data in which there are few zero values. Measures of this type include most information statistics (*i.e.* information analysis of Williams, Lambert and Lance, 1966) and the product moment correlation (as used by Coetsee

* The Bray and Curtis measure or Czekanowski coefficient or Sorenson coefficient (Bray and Curtis, 1957; Field, 1971; Williams *et al.*, 1973) is:

$$C = \frac{2w}{A + B} \times 100$$

where *w* is the sum of the lesser values of the species scores in the two plots compared; *A* and *B* are the sums of the species scores in each plot; *C* is the percentage similarity between the plots.

† The data used in Figs. 1 and 2 are from Campbell (1974) and consist of 39 relevés with 65 species. The relevés were collected from forest patches on the Cape Peninsula. The degree of inter-relevé variability or heterogeneity in the data was low; approximately 20% of the values in the data matrix being zero.

et al., 1973). This latter coefficient also suffers from another problem; its sensitivity to the departure of a species abundance value from the mean value in a plot (Hall, 1969).

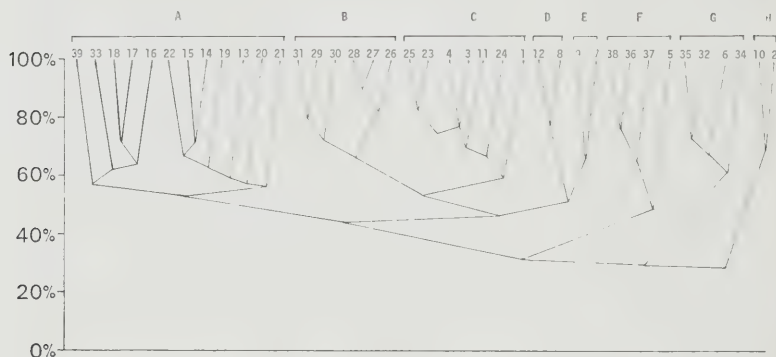


FIG. 1a.

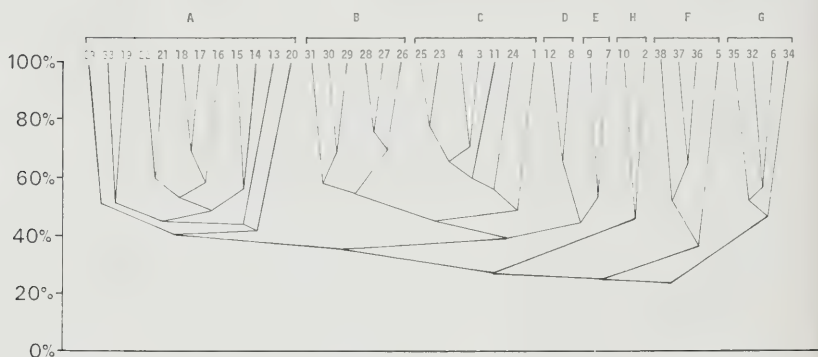


FIG. 1b.

Comparison of the group average relationships (see text) of forest relevés collected by Campbell (1974) as given by (a) the abundance weighted Bray and Curtis measure and (b) Canberra measure which is not abundance weighted. The letters A-H indicate the groups recognised. The scale shows percentage similarity.

The frequency modulated relative homogeneity function (Hqm) of Hall (1969) was also considered. Using this function weighting of species abundance values can be varied by the investigator. When no weighting is used double zero matches affect the analysis. No weighting is therefore unsuccessful when there are a large number of zero values in the data matrix, as is shown in the test data (Table 1). Seventy percent of the values in this matrix are zero values. The degree of inter-relevé variability or heterogeneity in the data is, therefore, high.

TABLE 1.
Test Data
plot numbers

species numbers										
	1	2	3	4	5	6	7	8	9	10
1	5	3	5	-	-	-	-	-	-	-
2	2	2	1	-	-	-	-	-	-	-
3	1	1	-	-	-	-	-	-	-	-
4	-	1	2	-	-	-	-	-	-	-
5	1	1	-	-	-	-	-	-	-	-
6	-	-	-	5	5	3	-	1	-	-
7	-	-	-	-	1	-	5	5	4	-
8	5	5	5	-	-	-	-	-	-	5

The classification of these data by Hqm (Fig. 2) shows that it is necessary to use Hqm fully abundance weighted if plot 10, which actually shows no similarity to plots 4, 5, 6, 7, 8 and 9, is not to be grouped with these plots. However, in other situations no abundance weighting may prove more successful than full abundance weighting. This can occur when the data matrix has few zero values, as was the case in the forest data used to construct Fig. 1. Here abundance weighting provided a classification that was difficult to interpret whereas Hqm, with no abundance weighting, provided a classification similar to those produced by the Canberra measure and Bray and Curtis measure (Fig. 3).^{*} Hqm must therefore be discarded as a possible standard all-purpose similarity measure; its efficiency being dependent on the investigator's

* For sorting the similarity values obtained from the Canberra measure and Bray and Curtis measure the method of group-average sorting was used (see following discussion), while the method of average member sorting of Hall (1969) was used with Hqm. The two different sorting methods could not explain the differences in the dendrograms as the methods are similar; average member sorting having a minimally better averaging technique (Hall, 1969; Field, 1971). However, this method requires much more computation time.

subjective assessment of the data and his subsequent choice of abundance weighting.

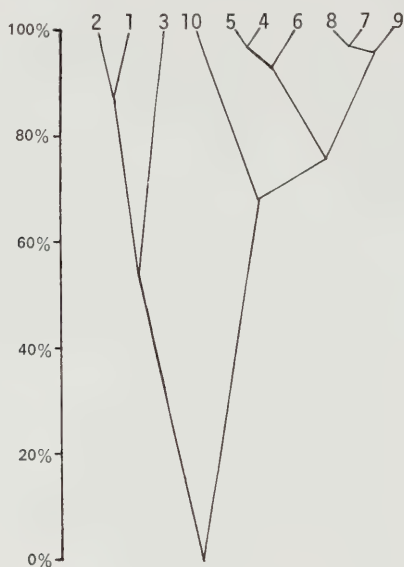


FIG. 2a.

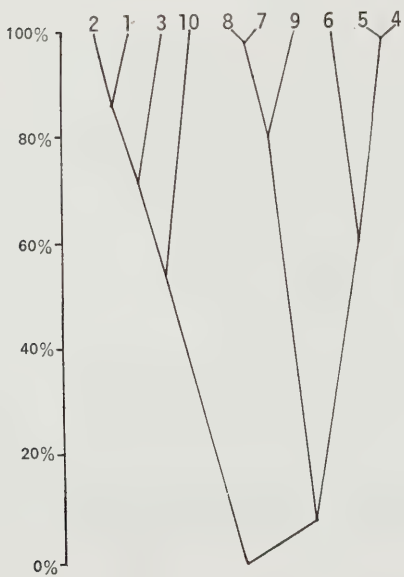


FIG. 2b.

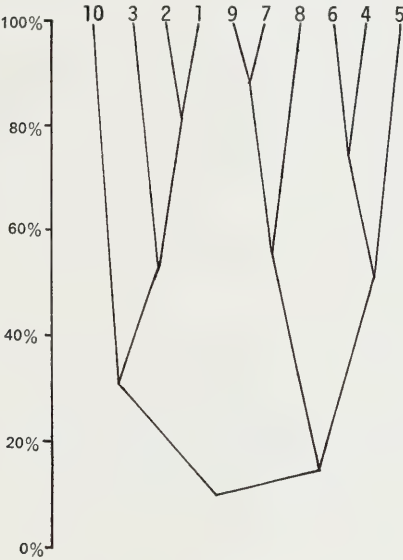


FIG. 2c.

Classifications of a raw data matrix with many zeros (Table 1) using (a) Hqm with 0% abundance weighting (b) Hqm with 100% abundance weighting and (c) the Canberra measure which excludes double zero matches. Plot 10 is misclassified in the first dendrogram.

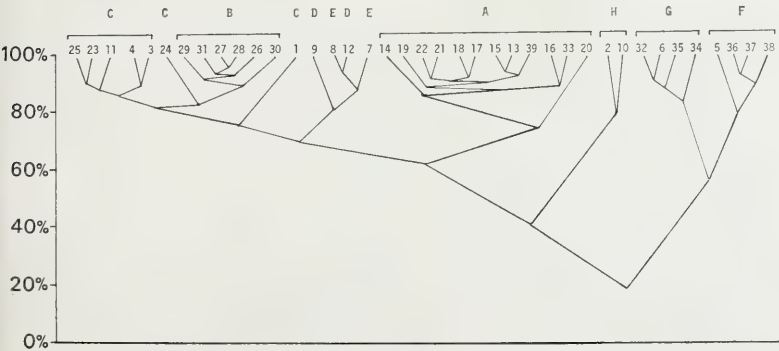


FIG. 3a.

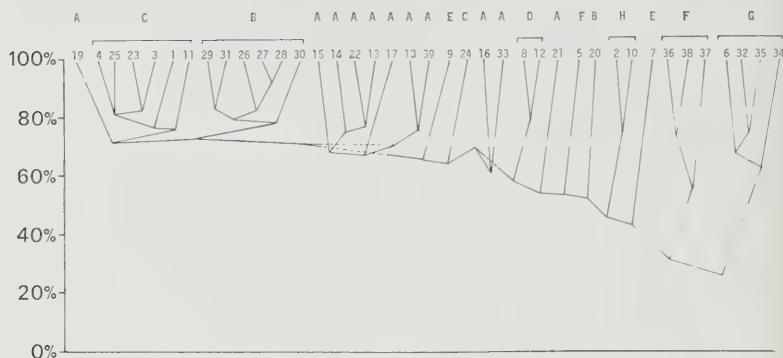


FIG. 3b.

Classifications of the forest data used for Fig. 1 (Campbell, 1974) using (a) Hqm with 0% abundance weighting and (b) Hqm with 100% abundance weighting. As opposed to Fig. 2 Hqm with 0% abundance weighting now provides more acceptable groupings. A-H indicate the groups that were recognised in Fig. 1.

One measure which would appear suitable is the Canberra measure.

$$C = \frac{1}{s} \sum_j \left\{ \frac{|x_{1j} - x_{2j}|}{x_{1j} + x_{2j}} \right\} \times 100$$

where x_1 is the abundance value of the j th species in plot 1; s is the total non double-zero matches of the species in the plots being compared, and C is the percentage dissimilarity of the plots. It has a slight problem in that when x_{1j} is zero the measure takes on its maximum value irrespective of the value of x_{2j} . This can be overcome by replacing zero by a small positive value for all zero/non-zero comparisons e.g. 0.1 (Williams *et al.*, 1973). This measure, being self-standardising over each comparison, is quite insensitive to large outlying values. It also does not consider double-zero matches (*cf* Fig. 2c). It would therefore appear to be of use as a similarity measure which can accept raw data which does not require manipulation prior to analysis; its success being independent of the nature of the raw data. It has produced favourable results in the work of Williams *et al.* (1973) and Webb, Tracey, Williams and Lance (1970), and also in forest and fynbos vegetation that the present authors are studying.

2. *Sorting (Cluster) technique*

The method of group-average sorting would appear to be suitable as a standard method for sorting or clustering items and groups of items. In studies by the present authors and by Field (1970, 1971) this method has proved successful. Hall (1969) mentions that information loss by this method is minimal. In this sorting method the most similar plots are grouped. The similarity coefficients between each plot of a newly formed group and a plot outside the group are then averaged. This is done for every plot outside a group to give new similarity coefficients between the group and the rest of the plots, so that in further groupings, the already formed groups are regarded as single plots.

Computation aspects

The time and effort required to prepare the raw data for computation may be cited as a problem in numerical analyses. However, in terms of time, the present authors have shown that to prepare a final Braun-Blanquet phytosociological table requires more time when manually prepared than when prepared with computer-based aids. We have prepared final phytosociological tables by using the vegetation classification as given by the Canberra measure and group-average sorting using a programme written by Dr. J. Field* and modified by B.C. Classification of species using the above technique has been unsuccessful. In tables completed to date the species have been subjectively arranged using a computer-based technique. As compared to manual methods of preparing phytosociological tables, both by rewriting and mechanical aids, the computer-based method has proved less time consuming and laborious, and is not a potential source of error. The data prepared for computation are then also available for analysis by a variety of numerical methods (*e.g.* ordination).

One serious problem with a computer-based method is the limitation on the size of the raw data matrix. This can however be overcome by methods such as those proposed by Janssen (1975). The essence of Janssen's method is that each relevé is considered separately, and only in relation to the relevés considered before. Clusters are formed through either assigning a new relevé to an already existing cluster, or to designate it as a separate cluster. In this way relevés can be divided into a number of clusters each of which can then be used in more detailed analyses.

CONCLUSIONS

Werger's (1974b) suggestion that the classification arrived at by numerical methods is not necessarily versatile, does not fully apply to the above technique.

* Zoology Department, University of Cape Town.

Forest data collected by Campbell (1974) were used to construct the dendrogram shown in Fig. 1. At a later date 66 further plots were collected from the forests under study and together with the initial plots were numerically analysed. The new data matrix now consisted of 105 plots with 120 species and a slightly higher inter-relevé variability; approximately 45% of the values being zero values. The resulting classification (Fig. 4) shows that the groupings obtained still hold in the final classification; except for plot 1 which, however, was the most atypical plot of group C. The method would therefore appear to be relatively versatile.

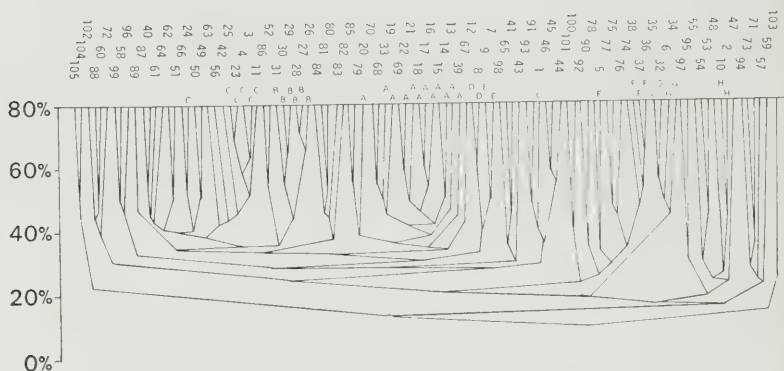


FIG. 4.

Group-average relationships of 105 forest relevés using the Canberra measure. Plots 1-39 were used in Fig. 1. A-H indicate the groups in which these plots were found in Fig. 1.

If a standard numerical method is available to classify relevés, the following benefits to phytosociologists can be envisaged:

1. a high reliability can be placed on the vegetation classifications obtained by different workers in different areas;
2. standardisation by different workers allows ease of comparison between workers;
3. a numerical method can provide an objective means of delineating syntaxa at the lower hierarchical levels where the differences between taxa are slight. This has been shown in our work on the Cape Peninsula where it has been possible to subdivide a large number of relevés with very similar floristics;
4. the classification which results can be expressed as a dendrogram which provides a clear graphical method of showing the similarity between relevés and groups of relevés;

5. an eventual aim of phytosociologists should be an inventory of the plant communities of South Africa. Computer-based methods are essential for this mammoth task. This aim will be made that much more attainable if the data are available in a machine readable form. Using a numerical method, quantitative values can be used to indicate syntaxonomical rank, e.g. 20–30% similarity may indicate the association level.

These benefits can only be realised if one of the number of equally recommended methods can be found suitable. The suggested method must avoid information loss and distortion and must provide a classification for general purposes. The results of the method should not depend on the nature of the raw matrix (*i.e.* raw table). As Goodall (1970) has stated, it is doubtful whether any single method will be consistently preferable. Different methods will stress different facets of the data. Nevertheless, group-average sorting based on similarity values obtained by the Canberra measure appears to be highly suitable as a general method of classifying relevés.

REFERENCES:

- BRAY, R. J. and CURTIS, T. T., 1957. An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* **22**: 235–349.
- CAMPBELL, B. M., 1974. *A Phytosociological survey of forest patches on Table Mountain.* University of Cape Town Honours project. Unpublished.
- ČEŠKA, A. and ROEMER, H., 1971. A computer program for identifying species-relevé groups in vegetation studies. *Vegetatio* **23**: 255–277.
- COETZEE, B. J., 1974. Improvement of association-analysis classification by Braun-Blanquet technique. *Bothalia* **11**: 365–367.
- COETZEE, B. J. and WERGER, M. J. A., 1973. On hierarchical syndrome analysis and the Zurich-Montpellier table method. *Bothalia* **11**: 159–164.
- FIELD, J. G., 1969. The use of information-statistic in the numerical classification of heterogeneous systems. *J. Ecol.* **57**: 565–569.
- FIELD, J. G., 1970. The use of numerical methods to determine benthic distribution patterns from dredgings in False Bay. *Trans. R. Soc. S. Afr.* **39**: 183–200.
- FIELD, J. G., 1971. A numerical analysis of changes in the soft-bottom fauna along a transect across False Bay, South Africa. *J. exp. mar. Biol. Ecol.* **7**: 215–253.
- GOODALL, D. W., 1970. Statistical plant ecology. *Ann. Rev. Ecol. Syst.* **1**: 99–124.
- GOWER, J. C., 1967. A comparison of some methods of cluster analyses. *Biometrics* **23**: 623–637.
- HALL, A. V., 1969. Avoiding information distortion in automatic grouping programmes. *Syst. Zool.* **18**: 318–329.
- JANSSEN, J. G. M., 1975. A simple clustering procedure for preliminary classification of very large sets of phytosociological relevés. *Vegetatio* **30**: 67–71.
- SHIMWELL, D. W., 1971. *Description and classification of vegetation.* London: Sidgwick and Jackson.
- WALKER, B. H., 1974. Some problems arising from the preliminary manipulation of plant ecological data for subsequent numerical analysis. *Jl S. Afr. Bot.* **40**: 1–13.
- WEBB, L. J., TRACEY, J. G., WILLIAMS, W. T. and LANCE, G. N., 1970. Studies in the numerical analysis of complex rain-forest communities. V. A comparison of the properties of floristic and physiognomic-structural data. *J. Ecol.* **58**: 203–232.
- WERGER, M. J. A., 1974a. On concepts and techniques applied in the Zurich-Montpellier method of vegetation survey. *Bothalia* **11**: 309–323.

- WERGER, M. J. A., 1974b. The place of the Zurich-Montpellier method in vegetation science. *Folia. Geobot. Phytotax. Praha.* **9**: 99-109.
- WILLIAMS, W. T., 1971. Principles of clustering. *Annual Review of Ecology and Systematics.* **2**: 303-326.
- WILLIAMS, W. T. and LAMBERT, J. M., 1959. Multivariate methods in plant ecology. I. Association-analysis in plant communities. *J. Ecol.* **47**: 83-101.
- WILLIAMS, W. T. and LAMBERT, J. M., 1961. Multivariate methods in plant ecology. III. Inverse association-analysis. *J. Ecol.* **49**: 717-729.
- WILLIAMS W. T., LAMBERT, J. M. and LANCE, G. N., 1966. Multivariate methods in plant ecology. V. Similarity analyses and information analyses. *J. Ecol.* **54**: 427-445.
- WILLIAMS, W. T., LANCE, G. N., WEBB, L. J. and TRACEY, J. G., 1973. Studies in the numerical analysis of complex rain-forest communities. VI. Models for the classification of quantitative data. *J. Ecol.* **61**: 47-69.